Novel approach Predicting Metal Abundances from Spatial Transcriptomics in Amyotrophic Lateral Sclerosis (ALS) and Gene Expression Analysis for Metal Abundances

Arav Srivastava

Dartmouth Health

ABSTRACT

- Amyotrophic Lateral Sclerosis (ALS)– Rare fatal neurodegenerative disease affecting 4-6 out of every 100,000 people. No cure or test.
- Excess of certain metals have relation to higher risk of ALS due to oxidative stress. High levels of heavy metals have been found in ALS patients.
- Gene expression and Metal Abundances have been studied in isolation never their relation
- Prediction of Metal Abundances from Gene expression and Gene Biomarkers from Metal abundances

INTRODUCTION

- **Metals relation to ALS:** Metals play a crucial role but increased levels of metals can sometimes be a cause for ALS due to oxidative stress caused by the metals. Excess metals can also disrupt the TDP-43 homeostasis. Some metals which in previous studies have been related with this are Copper, Iron, Selenium, Manganese, Zinc.
- **Gene expression relation to ALS:** Gene expression has also been studied in isolation similiar to metals and many studies have related genes like SOD1, TARDP, FUS, and C9orf72 directly to ALS. But the connection to these gene expressions to metal abundances has been an area which is poorly understood.
- **Spatial Transcriptomics (ST) and Elemental Imaging (EI):** new recent technology has been used in the creation of creating high-resolution elemental mapping like laser ablation, inductively coupled plasma mass spectrometry (LA-ICP-MS) and synchrotron-based X-ray fluorescence (XRF) microscopy [14]. These tools enable precise spatially resolved quantifiable metal abundance data which is extremely useful and a key for the future study of metal abundances

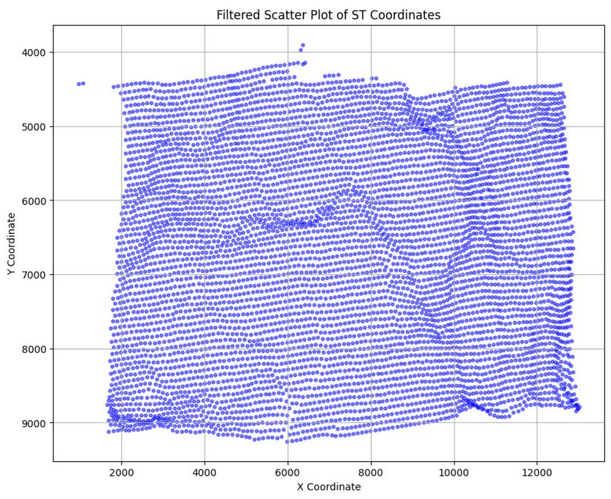

Figure 1: Spatial Transcriptomics coordinates for patient 021-13

METHODS

- **Goal:** Investigate the correlation between metal abundances and gene expression to make novel breakthroughs in ALS pathology
- **Data Gathering and Preprocessing:**
  - 10x Genomics Visium Spatial Gene Expression data (ST). Data of 4 patients with all 5 metals for abundance data (Fe65, Cu63, Zn66, Mn55, Mg24). Data is co-registered and aligned.
  - Filteration of NaNs, negative values, RobustScaler, StandardScaler used to make values more digestible for the model, log transformation, and choosing of 3000 highly variable genes for each of the patients out of the 18,085 for each patient to make training more efficient.
- **Model Building - Multilayer Perceptron (MLP) model:**
  - Domain adaptive MLP model created to train the model to predict metal abundances from gene expression data. Model consisted of multiple feature extraction and regression layers to train the model to learn biological relationships within the data.
- **Gene Expression Analysis:**
  - Combines approach of choosing 2000 highly variable genes for efficiency out of the 18,085 per patient and uses random forest regression. Correlation of each gene-metal pair was recorded. Model was trained with data from all 4 patients.
- **Model running and evaluation:**
  - MLP models R^2, RMSE, MSE, Accuracy, and Spearmen correlation coefficient were recorded. Model results on training set were used to determine model performance although Train, Test, and Validation metrics were all recorded.
  - 60/20/20 Train/Validation/Test split

RESULTS

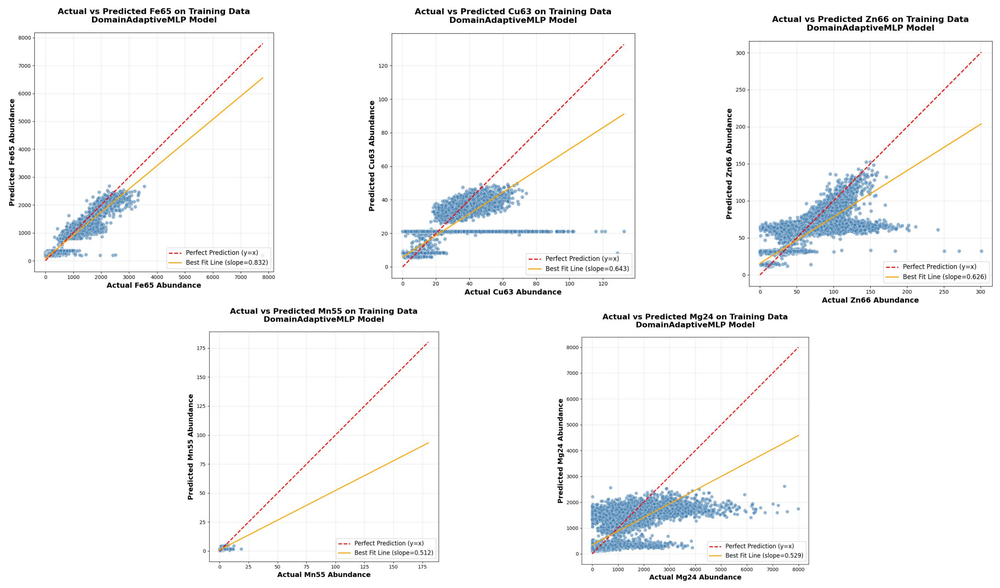

Figure 2: Line of best for true vs predicted model results in all 5 metals.

- Iron model (Fe65) performed the best showing the highest correlation with gene expression with an R^2 of 0.8984, MAE of 94.59, RMSE of 149.89, and Pearson correlation of 0.9510
- Copper model (Cu63) also performed well showing high correlation with gene expression with an R^2 of 0.7167, MAE of 5.15, RMSE of 8.65, and correlation of 0.8549
- Zinc model (Zn66) also performed well showing moderately high correlation with gene expression with an R^2 of 0.6965, MAE of 10.54, RMSE of 17.71, and correlation of 0.8413.
- Magnesium model (Mg24) showed more moderate performance with some but weak correlation most likely due to complex correlation to gene expression with an R^2 = 0.5962, MAE = 349.11, RMSE = 573.96, correlation = 0.7810.
- The Manganese model (Mn55) showed some correlation but not as strong as Copper, Iron, or Zinc with an R^2 = 0.5452, MAE = 0.4986, RMSE = 1.4534, and correlation of 0.7404.

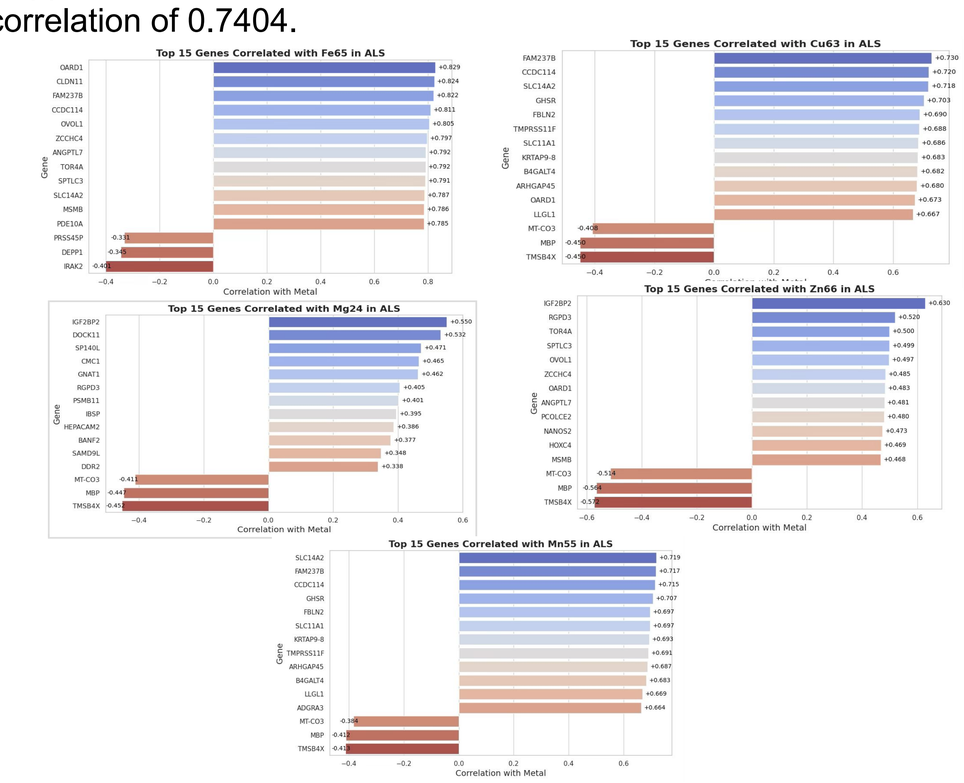

Figure 3: Top 15 gene correlations for each metal (Biomarker Analysis)

- Gene expression analysis revealed numerous genes associated with the certain metals. One gene which was prevalent across all metals was the FAM2378 gene which could possibly serve as a biomarker in ALS pathology along with many other genes

CONCLUSION

- **Potential for Clinical Impact:**
  - The model shows multiple new potential biomarker genes and proves the relationship between metals and gene expression in ALS. The novel study contributes to the understanding of ALS pathology within the scientific community and also provide possible new biomarkers in ALS pathology.
- **Limitations:**
  - The gene expression analysis was not thorough and lots of the genes could be normal genes which are common in the brain stem and not ALS-specific.
  - 4 patients doesn't provide a broad enough dataset to generalize across all patients in ALS although it gives a good insight into ALS pathology.
- **Future work:**
  - Patient based hold out split instead of a concatenated split. Further gene expression analysis to determine true biomarkers found.

Acknowledgements: EDIT, Dartmouth and Hitchock Medical Center, Aruesha Srivastava, Dr. Levy, Rishy Deosthali